[21] Data Collection Strategy

By Z. Dauter

The best way to proceed during X-ray diffraction data collection depends on qualitative factors, such as crystal quality and availability, type of X-ray source and detector, and time available, and quantitative ones, such as cell parameters, resolution limit, and crystal symmetry. There are certain rules to help one in producing a data set that is complete and accurate, and extends to as high resolution as possible. Often it is impossible to satisfy all these requirements simultaneously, and in most cases the actual data set collected is the result of a compromise. It is worth remembering that all subsequent steps of crystal structure analysis depend on the quality of data collected in the first instance; phasing, Fourier map interpretation, and refinement will proceed more smoothly if the data are good. To help define the parameters to use to set up data collection, in this chapter we discuss determination of the outer resolution limit, a precise description of the behavior of the reciprocal lattice during rotation photography, the effect of crystal mosaicity, the myth of the blind region, and finding ways in which crystal symmetry can help.

The least well-defined criterion in data collection is perhaps the resolution limit of diffraction. In principle, as long as the ratio of average intensity to the associated estimated error is higher than 1.0, the data contain some information. However, there may be only a few reflections having meaningful intensity among many reflections weaker than their associated errors. Therefore, extending the resolution limit may effectively introduce more noise than signal to the system, whether it is the Fourier transform or a least-squares matrix. A useful rule is to restrict the resolution to the point below which more than half of the intensities are higher than 2σ. This assumes that the errors of the measured intensities are estimated correctly. In most programs used for intensity integration from 2-D detector images,

Copyright © 1997 by Academic Press, Inc.
All rights of reproduction in any form reserved.

intensity uncertainties (σ) are estimated from counting statistics. However, these detectors do not count individual X-ray quanta, but produce a measure of intensity proportional to the number of recorded X-ray quanta. This proportionality is taken into account when uncertainties are estimated. Sigmas also can be estimated during scaling and merging from the spread of intensities of equivalent reflections: they can be put to the statistically expected value on the basis of the t plot, which tests the expected scatter of observables against the actual one as a function of sigma.

R_{symm} is less appropriate than I/σ as a criterion to judge the resolution limit because it depends on the multiplicity of the data and on the symmetry of the crystal. The lower the symmetry, the lower are the values of R_{symm}. A general rule of thumb is to accept all data up to R_{symm} (based on intensity) of about 30%, perhaps even higher for high-symmetry space groups, less for *P1*. For certain purposes, such as application of direct methods or Patterson-search techniques, the high-resolution limit may need to be extended. Here a small number of significant reflections is important; for direct methods only the strongest reflections are used anyway (those with highest normalized structure factors). In practice one could be generous in judging the high-resolution limit of diffraction at the data collection stage, and could make a decision later at the merging step, or even after preliminary refinement of the model. It is always possible to eliminate some useless data, but it is difficult to collect data again. However, it should be stressed that if the data quality is not satisfactory, it is better to collect them again than to try to use mediocre data, which would be a waste of time and effort.

The whole area of the detector window should be used. The crystal-to-detector distance should be as large as possible without losing resolution, because then the ratio of signal to noise is highest. The background spreads over a larger area of the detector, and diffracted rays are not so divergent.

To design a data collection protocol so that a data set is complete, i.e., (almost) all unique reflections are measured at least once, is not straightforward. Geometric considerations must be taken into account that depend on the crystal's orientation, its cell dimensions, and its symmetry. We first discuss the geometric principles of the rotation method and then individual cases of different symmetry of the crystal.

As the crystal is rotated during exposure of X-rays, the diffraction geometry may be explained by the Ewald-sphere construction, which in turn illustrates Bragg's law in three dimensions. The Ewald sphere with the radius $1/\lambda$ represents the radiation, and therefore is stationary, whereas the crystal is represented by the reciprocal lattice with the origin at the point where the primary beam leaves the sphere. The crystal is rotated

$$\frac{1}{2d} = \frac{1}{\lambda} \sin \vartheta \qquad\qquad \lambda = 2\,d\,\sin \vartheta$$

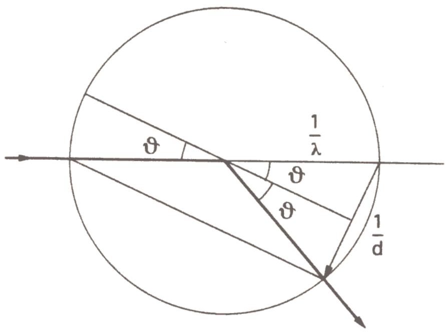

FIG. 1. Diffraction geometry using the Ewald sphere.

around one axis, usually perpendicular to the beam. During crystal rotation the reciprocal lattice points (rlps) cross the surface of the Ewald sphere, giving rise to diffraction, since the Bragg's diffraction condition is fulfilled (Fig. 1).

The rlps form the lattice, and they lie on planes in reciprocal space. If a plane, densely populated by rlps, is nearly perpendicular to the direct

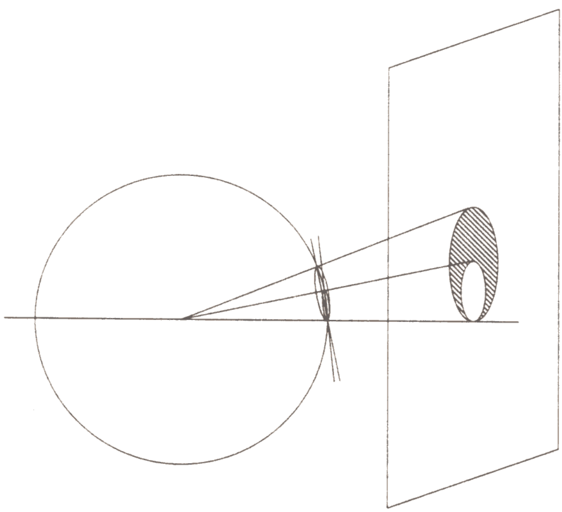

FIG. 2. Reciprocal lattice plane.

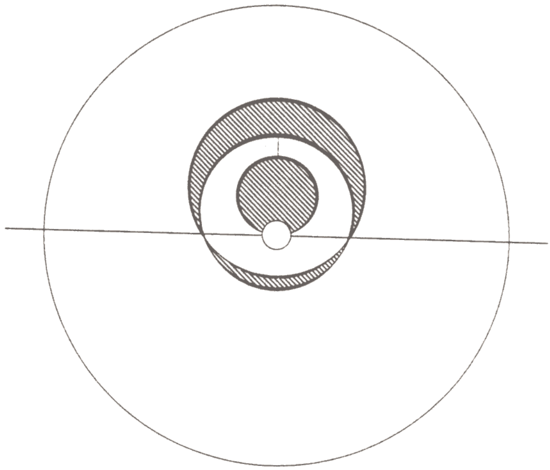

FIG. 3. Partially recorded reflections.

beam, it produces a cone of diffracted beams, since the cross section of the reciprocal-lattice plane and the Ewald sphere forms a circle. If the crystal does not move during the exposure of a so-called still image, it consists of a set of rings (ellipses) of spots, each ellipse arising from reflections from the same family of planes in the reciprocal lattice. If the crystal rotates

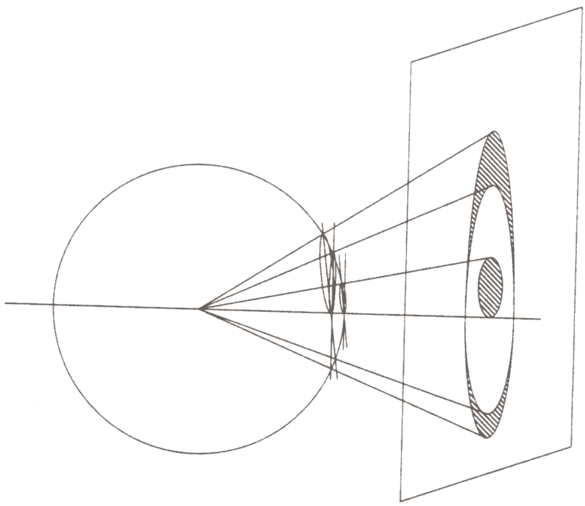

FIG. 4. Gap between two successive zones on detector.

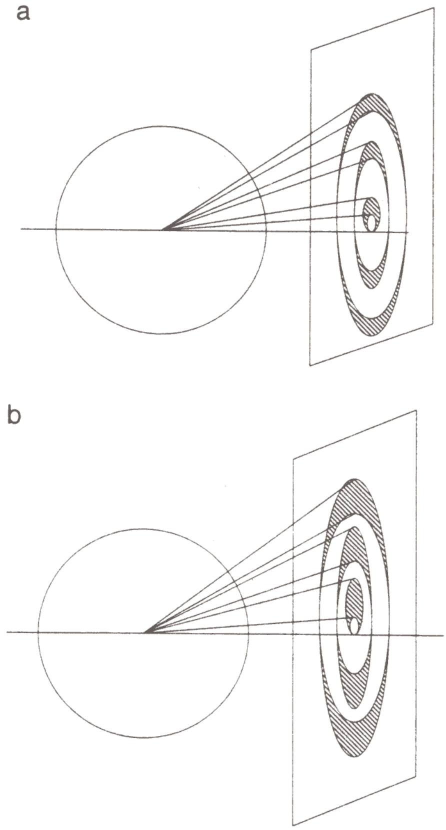

FIG. 5. (a–d) Overlapping of zones.

during exposure, the diffraction pattern will show all reflections in the area between two rings, corresponding to the start and end positions of the rotation (Fig. 2). If such a plane corresponds to the principal zone of the crystal, then all reflections of the ring will have one common index. If the zone passes through the center of the detector, it must correspond to the plane passing through the origin of the reciprocal lattice, as in Fig. 2. Such a principal zone passing through the center of the detector contains reflections with one index equal to zero.

At the ends of a rotation range, reflections at the edge of every zone will have started diffracting, but will not have passed completely through the surface of the Ewald sphere (Fig. 3). The remaining part of their intensity is then recorded on the next image. Such partially recorded reflections are called "partials," in contrast to reflections fully recorded on a single image.

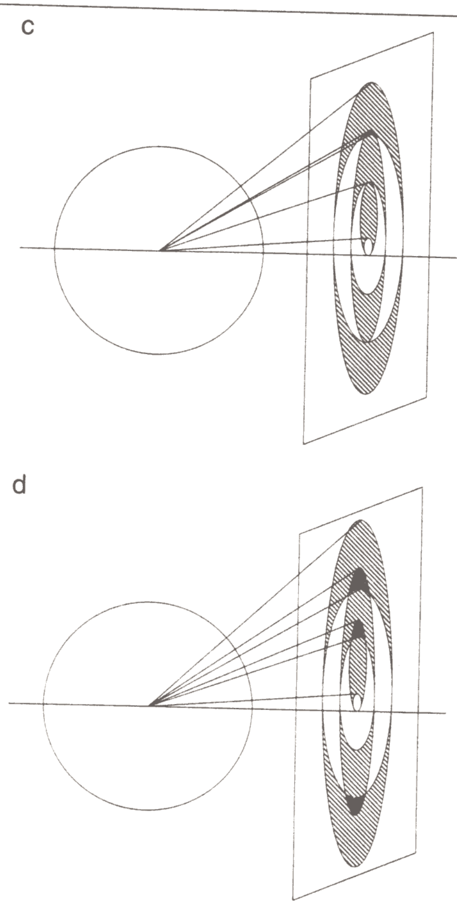

FIG. 5. (*continued*)

Clearly, if the effective rocking curve width (total effect of beam divergence and crystal mosaicity) is larger than the rotation range, all reflection are partials and there are no fully recorded reflections.

The gap between two successive zones on the detector depends only on the spacing between the two planes in the reciprocal lattice, i.e., the cell dimension in the direct lattice of the crystal (Fig. 4). The width of the zone on the detector plane (in the direction perpendicular to the rotation axis) is proportional to the amount of the crystal rotation during exposure. If the rotation increases, the zones become wider and at some point they will start overlapping, (Fig. 5 a–d). This effect will be visible first at high resolution at the edges of the image. Thus the permissible rotation per image

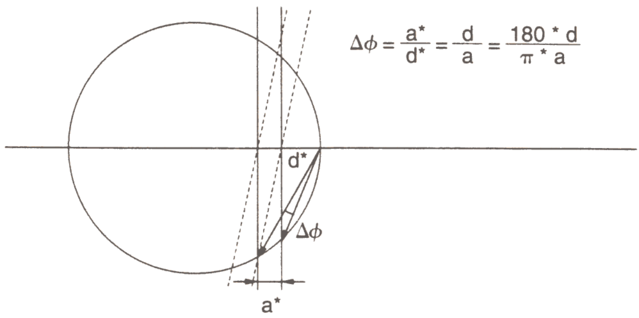

FIG. 6. Permissible rotation depends on resolution limit and cell dimension.

without overlap of the successive zones depends only on the resolution limit and cell dimension along the beam direction (Fig. 6). After conversion of radians to degrees it can be estimated from the formula

$$\Delta\phi = (180/\pi)(d/a)$$

where d is the resolution limit and a is the cell dimension. It must be stressed that this formula is only approximate. A few further remarks should be made in this context. The formula does not take into account the effect of crystal mosaicity, which is discussed further; the resulting rotation range should be diminished by the sum of crystal mosaicity and beam divergence (both effects make the zones wider).

The cell dimension in consideration is that along the beam in the given crystal orientation during exposure. As the orientation of the crystal changes from image to image, the maximum allowed rotation range may also change, since crystals often have different cell dimensions in different directions. However, the cell dimension along the spindle axis will never come into play, since the crystal is rotated around this axis and it will never lie along the beam. Therefore, in the case of a crystal with one cell dimension much longer than two others, it is most favorable to have it aligned on the goniometer along its axis. Unfortunately, crystals often grow slowest along the direction of the longest axis and often it is their thinnest physical dimension; it is difficult to mount the plate-like crystal across the capillary. However, for the cryoexperiment the loop can be bent to accommodate the flat crystals in the desired orientation.

If the two cell dimensions of the crystal that have a chance to be parallel to the beam (those not along the spindle axis) are significantly different, it is possible to use different rotation ranges in different orientations. However, the gain is not very high, since the maximal cell dimension requires small rotations over a wide range. For example, with an orthorhombic crystal having cell dimensions 100 and 50 Å, it may be necessary to use rotation ranges of 1° for about 60°, then perhaps 1.5° for the next 15°, and finally 2° for the last 15° of rotation when the short axis is nearly parallel to the beam direction. It may be noted here that it is much easier to make such decisions after one indexes the first image and simulates the real conditions with appropriate software.

The effect of the overlap on the detector of successive zones of reflections is important if the density of spots within the zone is high. If the crystal has small cell dimensions and the reflections are rather sparsely distributed on the detector (as for small-molecule crystals) this effect is not so important. The zones may be then allowed to overlap without the danger of overlapping reflection profiles (Fig. 7). The same is true if the crystal cell is centered: the reflections of one layer may be positioned between those of the previous or next zone. This is also true for skewed crystal orientations, when neither of the principal cell dimensions is oriented along the beam. As stated before, in such cases it is best to use the data processing software to find the crystal cell and orientation, estimate its mosaicity, and so on, and then to simulate the diffraction pattern for different orientations and rotation angles.

The mosaicity of the crystal can be judged easily by eye from the appearance of the zones. If they have very sharp edges with the reflection profiles of partially recorded reflections cut sharply, the crystal mosaicity is low. In contrast, if the zones do not have well-defined edges and reflection intensities fade out gradually at the zone border, the crystal mosaicity must

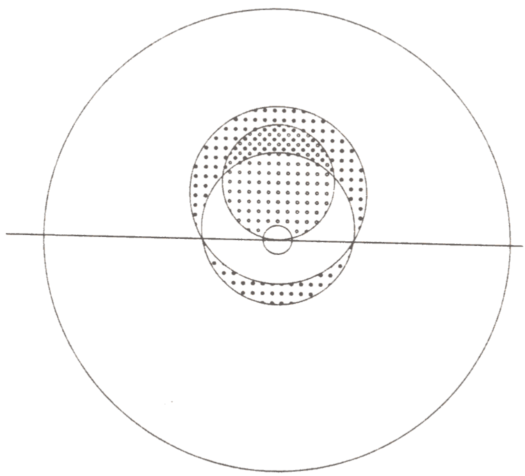

FIG. 7. Overlapping and density of spots.

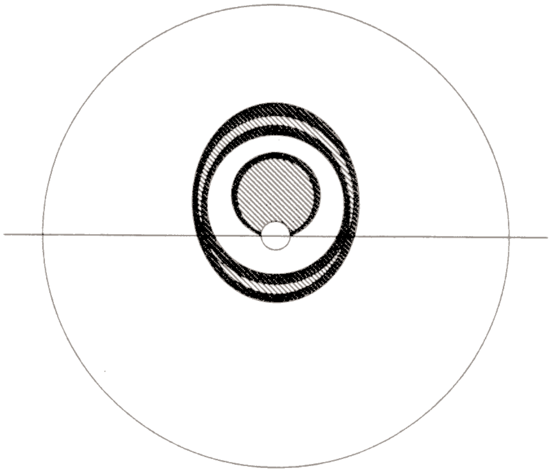

FIG. 8. Crystal mosaicity.

be high. Crystal mosaicity may also be estimated by the width of zones along the rotation axis: in theory in this direction the zones should have zero width on the detector. If they are wide along the central line parallel to the spindle axis, the crystal mosaicity is high (Fig. 8). Some crystals have anisotropic mosaicity, and its value depends on the crystal orientation.

It is possible to estimate the value of the mosaicity by simulating a predicted pattern and comparing it with the observed reflections present on the image. A more accurate value of mosaicity can be estimated by postrefinement after initial integration. If necessary, the integration should be repeated with use of the correct crystal mosaicity. Often highly mosaic crystals give quite good data; however, one must take this effect into account in estimating the allowed rotation range. It is better to use slightly too high a value of mosaicity during integration than too low. If it is underestimated, intensities of some partials are not integrated at all and are lost. If mosaicity is overestimated, the integration measures some nonexisting spots, but no intensities are lost, and after the summing of the two partial contributions, the total reflection intensity is preserved.

Another effect that is important for completeness of the data is the blind region, sometimes called the cusp. This is the region of the reciprocal space near the rotation axis, where reciprocal lattice points do not cross the surface of the Ewald sphere even after full 360° rotation. To illustrate this it is easier to imagine that the crystal (reciprocal lattice) is stationary and the Ewald sphere rotates (Fig. 9). After 360° of such rotation around the axis that is tangent to the sphere, all reflections that are within a torus

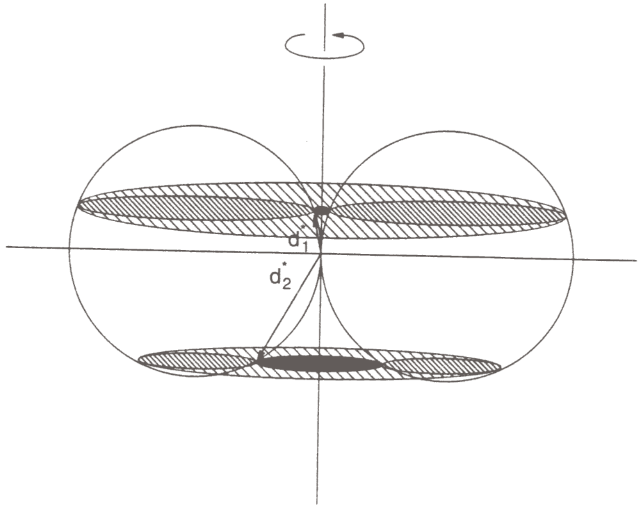

FIG. 9. Blind region.

centered at the origin have crossed the sphere surface twice, coming inside and going outside. Inner regions close to the axis on both sides of the torus constitute the blind region, containing rlps that will never be in a diffracting position, and therefore are impossible to be collected with the given crystal mount. The half-width of the blind region at the highest resolution is equal to the maximum θ angle and the ratio of its volume to the volume of the whole resolution sphere (Fig. 10):

$$B = 1 - \tfrac{3}{32}(4\theta - \sin 4\theta)/\sin^3 \theta$$

At low resolution, i.e., closer to the origin, the blind region is narrow and becomes wider as resolution increases (Fig. 11). Also for a short wavelength, when the Ewald-sphere radius is large, its surface is flatter, and consequently the blind region is smaller (Fig. 12); this is one of the advantages of using short-wavelength radiation. With a wavelength shorter than 1 Å the blind region does not exceed 5% even at very high resolution and at 2 Å resolution remains below 2%.

If the crystal has no symmetry (triclinic) there is no way to achieve 100% completeness from a single rotation pass during data collection, no matter what the crystal orientation is. Even assuming that only one hemisphere of the reciprocal space is necessary because of the Friedel's law, the upper part of the blind region is related to the bottom part by a center of symmetry; therefore, both Friedels are lost. In such a case the only remedy is to remount the crystal to make it rotate around a different

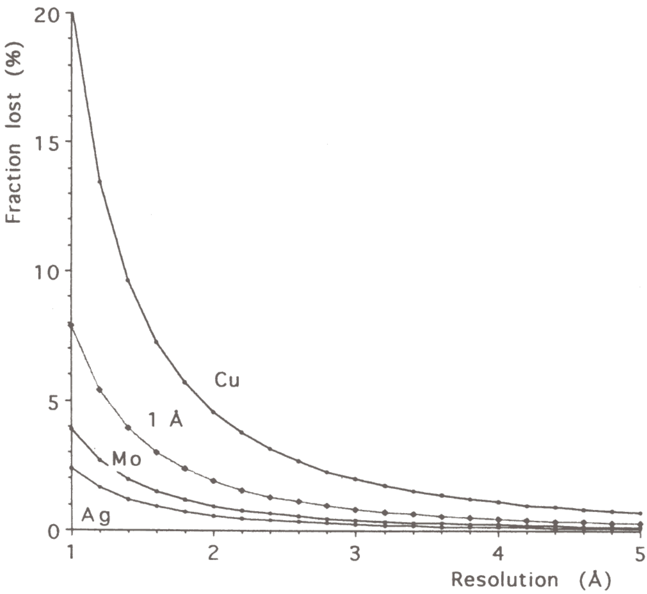

FIG. 10. Fraction lost versus resolution.

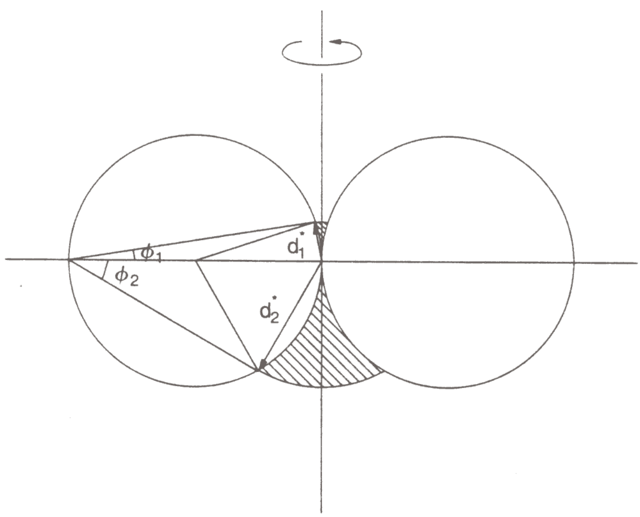

FIG. 11. Blind region at different resolutions.

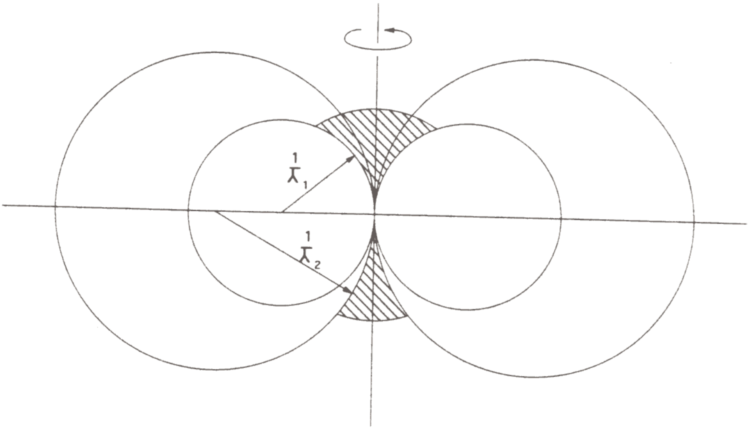

FIG. 12. Blind region affected by short wavelength.

direction and to cover the missing reflections in the second pass. The rotation axis has to be skewed by an angle larger than 2θ at the highest resolution. Usually this can be achieved using only the goniometer arcs.

If the crystal has symmetry higher than triclinic, it should always be possible to orient it in such a way that reflections within the blind region will have their symmetry equivalents further away from the rotation axis, outside of the blind region, and therefore possible to be measured (Fig. 13). Only if the crystal is oriented along one of its unique axes will the blind region affect the completeness of data. It is therefore advisable to skew the crystal from the ideal axial orientation by at least the value of θ_{max}. Only in some special cases may it be advantageous to have the crystal perfectly aligned along its symmetry axis: for example, for measuring accurate anomalous dispersion data, where it is good to have both Friedel mates present on the same image.

The most important parameters to govern the completeness of data are the crystal symmetry and the choice of the proper rotation start and end position. In the best case, one should collect 180° of data to give a highly redundant set of reflections for accurate merging. Often this is impractical, so only the asymmetric part of the reciprocal space is covered; it is important to cover the correct part.

Looking along the rotation axis one can see that rotation by 180° is enough to collect all reflections within a hemisphere (Fig. 14). Some of them will be collected two times, by having crossed both the upper and lower part of Ewald sphere surface. Therefore, for a triclinic crystal, 180° rotation is enough to have all of the asymmetric unit collected (apart from

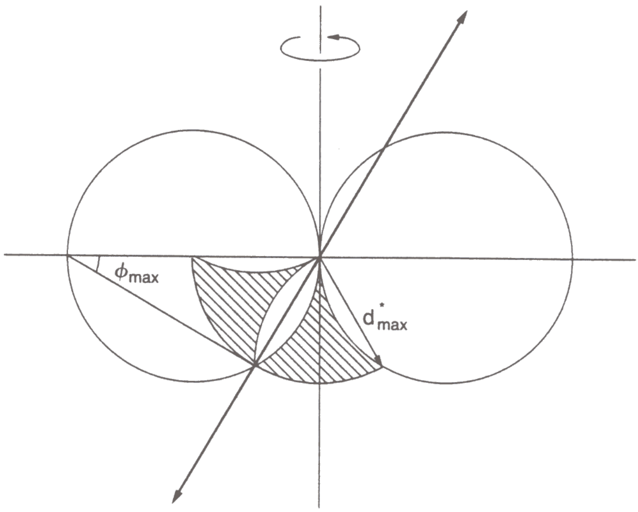

FIG. 13. Crystal orientation affecting blind region.

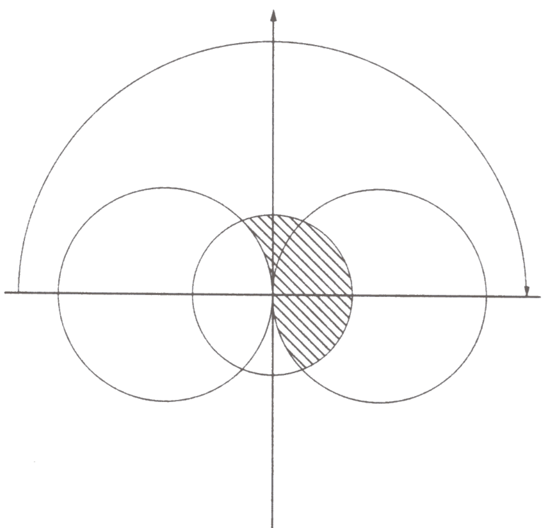

FIG. 14. Collection of reflections within a hemisphere.

the blind region). This is true if Friedel's law is assumed; in an anomalous-scattering experiment more than 180° are required to complete the data, but not necessarily the whole 360°. This results from the Ewald sphere being curved. If it were flat, 180° would suffice in all cases. In a low-resolution anomalous-scattering experiment on a P1 crystal, little more than 180° is required; at high resolution more is needed.

A similar situation occurs when a monoclinic crystal is rotated around its twofold axis. The asymmetric part consists of 180° of rotation, irrespective of the start position. For anomalous diffraction, more than 180° are not necessary since reflections related by the twofold axis are equivalent and are not Friedel related. Bijvoet mates are then collected on the left and right sides of the detector on the same image, since they are related by the mirror perpendicular to the twofold axis. It is worth noting that because of the curvature of the Ewald sphere, the completeness is not directly proportional to the fraction of the total required rotation range. After 90° of rotation when 180° are required, the data are about 65% complete (Fig. 15). However, when the total rotation approaches 180°, the high-resolution part of the data may be completed already, and at the end only the low-resolution part is filled up (Fig. 16). If data are collected in two or more passes, with long exposures at high resolution and shorter exposures at lower resolution (to measure strong, previously overloaded reflections),

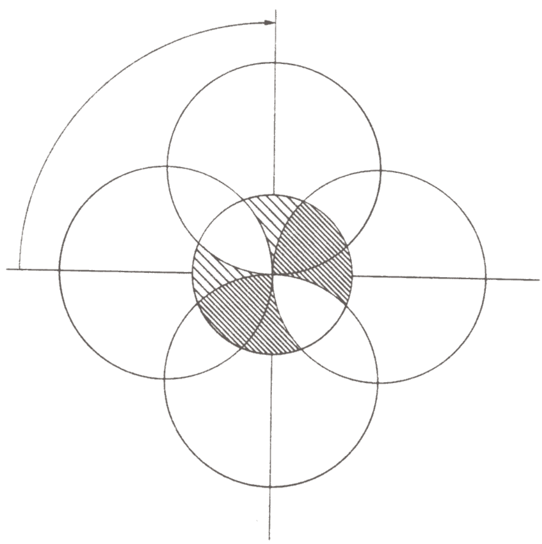

FIG. 15. Rotation of monoclinic crystal around twofold axis.

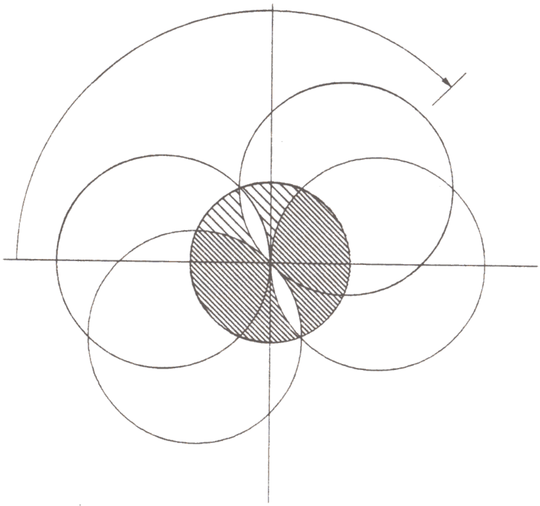

FIG. 16. Completeness at low and high resolution.

the first pass does not need to be fully completed, but the lowest resolution pass must cover all the required rotation range.

If a monoclinic crystal is mounted along an axis perpendicular to its twofold axis, i.e., the spindle axis direction lies in the x, z-plane, the required asymmetric part corresponds to 90° of rotation. However, in this case it is important that this be between the twofold axis and the x, z-plane. In other words, the start of data collection must be when the twofold axis is along the beam and end when it is in the detector plane, or vice versa. Note again that to avoid a blind region the crystal should be misoriented from either being along the twofold axis or lying in the monoclinic plane.

If an orthorhombic crystal is rotated around a direction close to one of its three mutually perpendicular twofold axes, 90° of rotation are sufficient for a complete data set (Fig. 17). It must be the right 90°: a 90° rotation between two diagonal positions results in ones collecting 45° of data twice and gives an approximately 65% complete set (Fig. 18). To avoid a blind region, the crystal should be misoriented; in the most advantageous orientation, the rotation axis should lie anywhere within one of the principal crystal planes.

Table I lists all noncentrosymmetric space groups according to the crystal class. In Table II, we summarize the required rotation range for crystals of different symmetry (crystal class) in different orientations, assuming the central, symmetric position of the detector with respect to the beam; if it is tilted by some value of 2θ, more rotation is necessary. For cubic

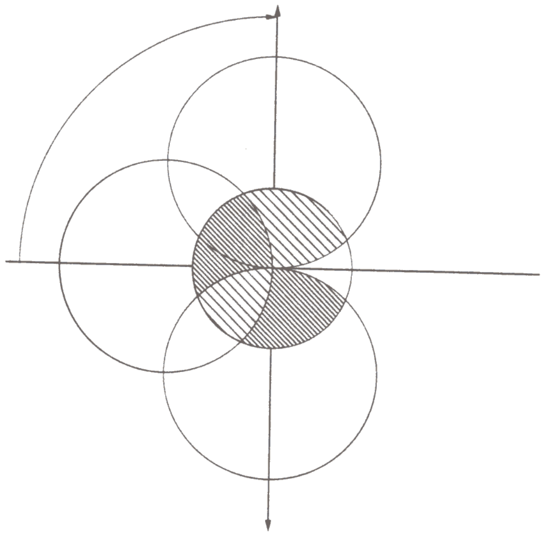

FIG. 17. Rotation of 90° for orthorhombic crystal.

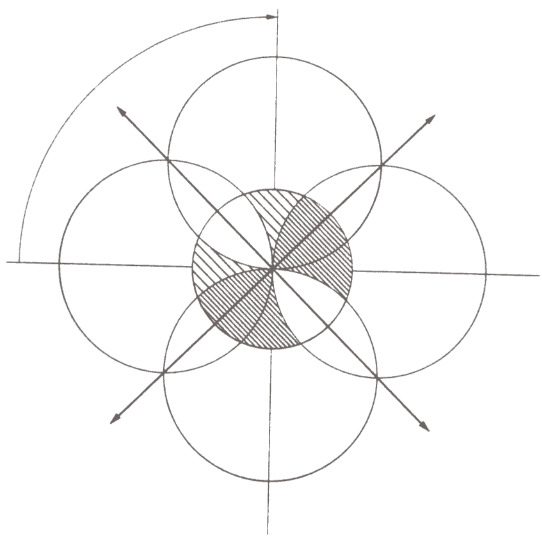

FIG. 18. A 90° rotation between two diagonal positions.

TABLE I
NONCENTROSYMMETRIC SPACE GROUPS

Crystal structure	Space group[a]
Triclinic	$P1$
Monoclinic	$P2$, $P2_1$, $C2$
Orthorhombic	$P222$, $P222_1$, $P2_12_12$, $P2_12_12_1$, $C222$, $C222_1$, ($I222$, $I2_12_12_1$), $F222$
Tetragonal	$P4$, ($P4_1$, $P4_3$), $P4_2$, $I4$, $I4_1$
	$P422$, ($P4_122$, $P4_322$), $P4_222$, $P42_12$, ($P4_12_12$, $P4_32_12$), $P4_22_12$, $I422$, $I4_122$
Trigonal	$P3$, ($P3_1$, $P3_2$), $R3$
	$P321$, ($P3_121$, $P3_221$), $R32$
	$P312$, ($P3_112$, $P3_212$)
Hexagonal	$P6$, ($P6_1$, $P6_5$), ($P6_2$, $P6_4$), $P6_3$
	$P622$, ($P6_122$, $P6_522$), ($P6_222$, $P6_422$), $P6_322$
Cubic	$P23$, $P2_13$, ($I23$, $I2_13$), $F23$
	$P432$, ($P4_132$, $P4_332$), $P4_232$, $I432$, $I4_132$, $F432$, $F4_132$

[a] In parentheses are groups indistinguishable by diffraction.

crystals it is difficult to give an exact estimate of necessary rotation. Certainly 60° for the class 23 and 35° for the class 432 will give complete and highly redundant data. Also, in the collection of anomalous data in higher symmetry groups, less rotation than theoretically required will give a high percentage of both Bijvoet mates.

If there is no time to rotate the crystal through the whole required

TABLE II
TOTAL ROTATION REQUIRED FOR COMPLETE DATA IN CASE OF SYMMETRIC DETECTOR POSITION

Crystal class	Point group	Rotation required for	
		Standard data	Anomalous data
Triclinic	*1*	180°	360°
Monoclinic	*2*	180° (b^*), 90° (a^*, c^*)	180° (a^*, b^*, c^*)
Orthorhombic	*222*	90° (a^*, b^*, c^*)	90° (a^*, b^*, c^*)
Tetragonal	*4*	90° (a^*, b^*, c^*)	90° (c^*), 180° (a^*, b^*)
	422	45° (c^*), 90° (a^*, b^*)	45° (c^*), 90° (a^*, b^*)
Trigonal	*3*	60° (c^*), 90° (a^*, b^*)	120° (c^*), 180° (a^*, b^*)
	321	30° (c^*), 90° (a^*, b^*)	60° (c^*), 180° (a^*, b^*)
	312	30° (c^*), 90° (a^*, b^*)	60° (c^*), 180° (a^*, b^*)
Hexagonal	*6*	60° (c^*), 90° (a^*, b^*)	60° (c^*), 180° (a^*, b^*)
	622	30° (c^*), 90° (a^*, b^*)	30° (c^*), 90° (a^*, b^*)
Cubic	*23*	About 60°	About 70°
	432	About 35°	About 45°

TABLE III

SPACE GROUPS WITH MORE THAN ONE INDEXING POSSIBILITY[a]

$P4$, ($P4_1$, 4_3), $P4_2$, $I4$, $I4_1$
$P3$, ($P3_1$, $P3_2$), $R3$
$P321$, ($P3_121$, $P3_221$), $P312$, ($P3_112$, $P3_212$)
$P6$, ($P6_1$, $P6_5$), ($P6_2$, $P6_4$), $P6_3$
$P23$, $P2_13$, ($I23$, $I2_13$), $F23$

[a] In parentheses are groups indistinguishable by diffraction.

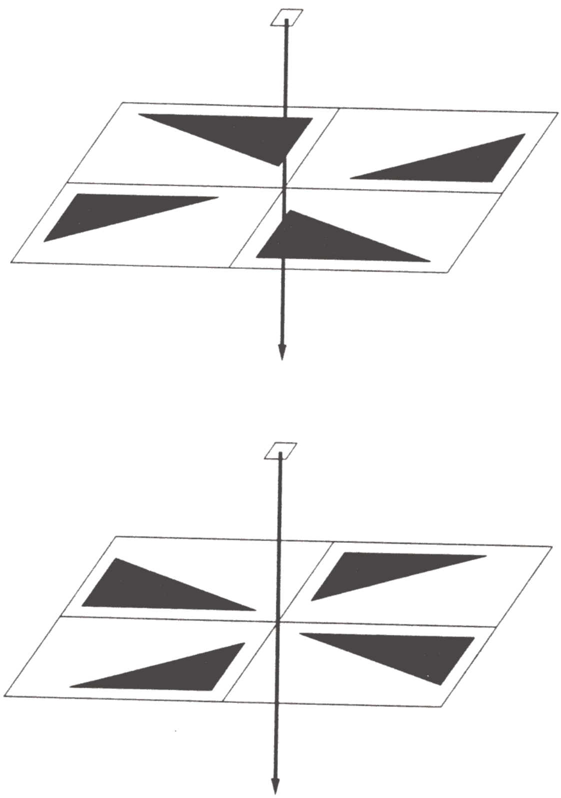

FIG. 19. Space groups with polar axis can be indexed up or down.

range, it is more effective to split the range into several slices and leave uncovered gaps between them. Three 20° regions separated by 10° gaps will give a higher overall completeness than a continuous 60° rotation and 30° left out of the required 90°. This effect also results from the curvature of the Ewald sphere.

In certain space groups a diffraction pattern can be indexed in more than one way (Table III). All the different indexing possibilities are valid, but they are not equivalent to one another. All space groups with a polar axis (classes 4, 3, and 6) can be indexed "up" or "down" this axis (Fig. 19). In cubic class 23 there is no fourfold axis, but the lattice is built up from perfect cubes; in this case rotation of 90° around any of the principal axes leads to another, nonequivalent indexing. This is important: when one collects data from more than one crystal or compares the native and derivative data, all reflections must be indexed the same way.

At the end it is worth remembering that it is advisable for one to spend some time interpreting the first image one records, either by eye or preferably with the data-processing software. One can adjust the conditions and simulate diffraction patterns at different crystal orientations to find an ideal strategy, rather than collecting data blindly and at some risk. You deserve to take the most from your crystals.

METHODS IN ENZYMOLOGY, VOL. 276

Copyright © 1997 by Academic Press, Inc.
All rights of reproduction in any form reserved.